Netherlands organization for applied scientific research

IZF

TNO Institute for Perception

IZF 1989-19

A. Toet

IMAGE FUSION THROUGH MULTI-RESOLUTION CONTRAST DECOMPOSITION

16

Netherlands
organization for
applied scientific
research

TNO-report

TNO Institute for Perception

P.O. Box 23
3769 ZG Soesterberg
Kampweg 5
3769 DE Soesterberg, The Netherlands
Fax +31 3463 5 39 77
Phone +31 3463 5 62 11

IZF 1989-19

A. Toet

IMAGE FUSION THROUGH MULTI-RESOLUTION CONTRAST DECOMPOSITION

16

Classifications:

Copies	: 50	Report	: Unclassified
Number of pages	: 34	Title	: Unclassified
		Abstract	: Unclassified

CONTENTS

Report No. : IZF 1989-19

Title : Image fusion through multi-resolution contrast decomposition

Author : Dr. A. Toet

Institute : TNO Institute for Perception
TNO Division of National Defence Research
Group: Vision

Date : April 1989

HDO Assignment No. : A87/D/149

No. Program of Work : 731.2

ABSTRACT

Integration of images from different sensing modalities can produce information that cannot be obtained by viewing the sensor outputs separately and consecutively. This report introduces a hierarchical image merging scheme based on multiresolution contrast decomposition. The composite images produced by this scheme preserve those details from the input images that are most relevant to visual perception. As an example the method is used to merge parallel registered thermal and visual images. The examples show that the fused images present a more detailed representation of the depicted scene. Detection, recognition and search tasks may therefore benefit from this new image representation.

Rap.nr. IZF 1989-19, Instituut voor Zintuigfysiologie TNO, Soesterberg

---

Beeldintegratie door Multi-Resolutie Contrast Ontbinding

A. Toet

SAMENVATTING

Dit rapport introduceert een methode om een enkel beeld samen te stellen uit een willekeurig aantal afzonderlijke beelden. De uitgangsbeelden mogen afkomstig zijn van verschillende typen beeldvormende systemen en mogen van resolutie verschillen. De enige restrictie van de methode is dat de uitgangsbeelden een zekere spatiële overlap moeten bezitten.

De methode werkt als volgt. Eerst worden de afzonderlijke beelden ontbonden in contrastrijke details van verschillende afmetingen. Vervolgens worden op overeenkomstige locaties in de verschillende beelden de details met het hoogste contrast geselecteerd. Tenslotte wordt het samengestelde beeld geconstrueerd uit de geselecteerde details. Het samengestelde beeld bevat juist die details van de afzonderlijke uitgangsbeelden die voor de visuele perceptie het meest relevant zijn. Ter illustratie worden de resultaten getoond die werden verkregen door de combinatie van warmte- en visuele beelden. De voorbeelden tonen dat de gecombineerde beelden een vollediger beschrijving geven van de afgebeelde scène. Door de toename van de informatie-inhoud van de samengestelde beelden lijkt het waarschijnlijk dat de methode kan bijdragen tot een verbeterde taakprestatie in uiteenlopende militaire omstandigheden.

1 INTRODUCTION

1.1 The benefits of sensor fusion

The problem of target detection, tracking and classification has been the subject of a multitude of theoretical and practical design activities for many years. While some surveillance systems use only a single sensor to implement these functions, the present trend is to incorporate multiple sensors.

Systems that use multiple sensors have many benefits. Generally they perform more functions and operate under a larger range of conditions than a single sensor. They show a graceful performance degradation in the event of component failures. As a result, they provide an increased resistance to countermeasures.

The use of multiple sensors of the same or of a different type provides an increased number of observations. This may result in redundant data, complementary data or both. Redundant data are obtained when two or more sensors measure the same parameters in an overlapping coverage. Complementary data are obtained if different parameters are measured and/or the coverage of a given area is not accomplished in an overlapping manner. Both cases lead to an increase (i.e. improvement) in processing gain. The degree and importance of this increase depend on the particular system application. Generally, overlapping coverage may yield improvements in target parameter accuracies, while data complementation can improve the detection sensitivity of the overall system.

In summary, the specific potential benefits from multisensor combination are:

- increased sensitivity
- higher accuracy and resolution
- enhanced target recognition.

1.2 The aim of image fusion

Nowadays there is a large number of imaging modalities in use (e.g. direct view optics, television, forward looking infrared, infrared search and track, microwave radar, millimeter wave radar, laser radar, synthetic aperture radar, laser rangefinder, acoustic transducer array, radio frequency interferometer etc.). Systems that use a number of imaging systems severely increase the workload of a human operator. Moreover, a human observer cannot reliably integrate visual information by viewing multiple images separately and consecutively.

The integration of information across multiple human operators is nearly impossible. An imaging system that fuses signals from multiple imaging sensors into a single image is therefore of great practical value.

Image fusion can be performed on different levels of abstraction. In general it is best to fuse image data at the lowest possible level, i.e. before information is lost in any thresholding or time sampling process. Such methods are applicable to multi-spectral infrared or visible light images in which the ratio of intensities of a corresponding pixel (picture element or sample) is a measure of the colour temperature. However, for pixel-level fusion there must be exact spatial registration between the sensors and preferably some dimensional similarity. In general, pixel-level fusion leads to high computation load on the processor and to a system that is neither modular nor fault tolerant.

Next in the hierarchy of image fusion methods is data fusion at the feature-level. Feature level fusion allows each sensor to provide quantitative information on the characteristics of its input at any instant and in any dimension. If the sensors are suitably registered the set of extracted features will correspond to some object in space. These features can simply be the result of (extracted by) convolution of the input function with some arbitrary filter function. An example of a feature that can be used in target verification is the local "pointlikeness" or "blobness" of an image intensity distribution. This measure can simply be extracted by convolving the image with a Laplacian filter. The spatial scale of the filter directly relates to the pronouncedness of the target. A "feature-vector" is obtained when a number of different features is derived for each region in space. Sensor fusion on the level of feature-vectors can for instance be obtained by the use of associative neural nets (Eklundh *et al.*, 1986; Pearson and Gelfand, 1988). These parallel distributed computational structures can efficiently classify multimodal vectors. Moreover, they may also incorporate contextual information.

Finally, sensor fusion can be obtained at a high level of abstraction involving hierarchical symbolic signal representations (O and Toet, 1989). The symbols in these descriptions represent structural primitives (i.e. meaningful features) of the signal. The hierarchical relation describes the structural composition of the signal. In general, the hierarchical signal representation is obtained by three consecutive or combined operations, namely application of a filter operation, extraction of the signal features, and determination of the hierarchical structure. At each hierarchical level, the filter operator removes unnecessary details in the image.

Thereafter, the signal representation is obtained by application of a suitable description method. Finally, the hierarchical relation determines the connection between signal representations at subsequent hierarchical levels. In hierarchical signal representations global information can be used to impose constraints on local operations. Thus, hierarchical operations can be more efficient than operations performed at a single level of abstraction. Moreover, global information may raise the confidence level of local estimates.

This report presents a scheme for the fusion of signals from multiple imaging systems. The input of the algorithm is an arbitrary number of simultaneously registered images. The images may be of different modalities and resolution. The only restriction of the method is that the input images must have some degree of spatial overlap. Each input image is decomposed into a set of perceptually relevant pattern primitives. Pattern sets for the various source images are then combined to form a single set for the composite image. Finally the composite image is reconstucted from its set of primitives. As a result the output of the algorithm is a composite image that preserves those details from the input images that are most ralevant to visual perception.

To illustrate the new image fusion scheme it is applied to fuse visual and thermal images.

2 HIERARCHICAL IMAGE FUSION

The essential problem in merging images for visual display is "pattern conservation": important details of the component images must be preserved in the resulting composite image while the merging process must not introduce spurious pattern elements that could interfere with subsequent analysis. Simple methods to combine image details often create edge artifacts between regions taken from different images (e.g. cutting and pasting, sometimes followed by edge blurring) or may annihilate image details (e.g. grayscale addition).

In this section we present a scheme to integrate signals from multiple imaging systems. The scheme employs a hierarchical symbolic representation of the input signals. The hierarchical representation is based on the fact that image structure depends on resolution.

2.1 The resolution dependency of image structure

When we zoom in on an image detail we clearly see its substructure but we will loose the articulation of its outlines. On the other hand, when we defocus an image to obtain a good overview of the depicted scene details of that scene will be hard to recognize because they are heavily blurred. It appears that relevant details of images exist only over a restricted range of the resolution parameter.

When resolution is decreased images become less articulated because the extrema (light and dark blobs) disappear one after the other. As a result any image can be represented by a juxtaposed and nested set of light and dark blobs, wherein each blob has a limited range of resolution in which it manifests itself (Koenderink, 1984; Toet *et al.*, 1984).

2.2 The construction of a multiresolution image representation

In the sequel we will refer to filter operators that eliminate details smaller than a certain size as *size limiting filters*. The difference of two size limiting filters will be called a *size selective filter*.

A family of images with progressively decreasing structural content (i.e. progressively less detail) can be produced by repeated application of a size limiting filter operator of a progressively increasing size limit.

A hierarchical image representation can be obtained by relating the descriptions generated by filters of increasing size limits. A trivial hierarchical relation is the one-to-one relation. In this case a correspondence is established between all points of successively filtered versions of the image.

Descriptions generated by large size filter operators contain no small size details. This consideration allows a progressive reduction of the sample frequency with a progressive increase of the filter size. Sampling induces a natural hierarchical relation.

A well known hierarchical image representation is the quad-tree or pyramid (Rosenfeld, 1984; see Fig. 1). A pyramid is a sequence of images in which each image is a filtered and subsampled copy of its predecessor. Successive levels of a pyramid are generally reduced resolution versions of the input image (hence the term "multiresolution" representation; e.g. Fig. 5). However, the representation may also consist of descriptive information about certain

image features (e.g. edges). In this case successive levels of the pyramid represent increasingly coarse approximations to these features (e.g. Fig. 6).

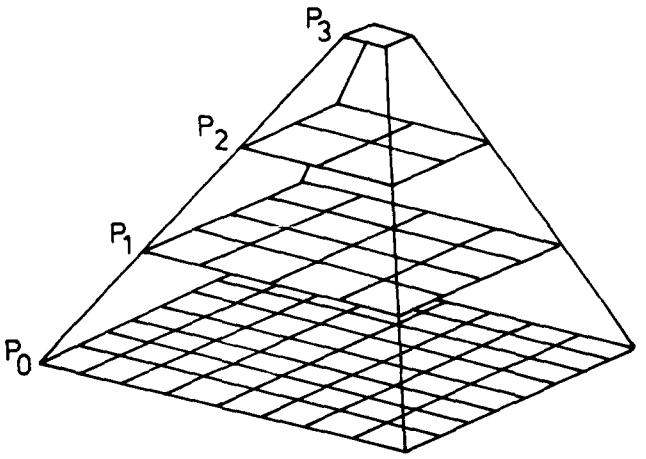

Fig. 1 Schematical representation of a pyramid data structure.

A pyramidal image representation is constructed in the following way. The original image is the bottom or zero level P_0 of the pyramid. Each node of pyramid level l ($1 \leq l \leq N$ where N is the index of the top level of the pyramid) is obtained by sampling a filtered version of level l - 1. The process which generates each image in the sequence from its predecessor will be called a REDUCE operation since both the sample density and the resolution are decreased. Thus for $1 \leq l \leq N$ we have

$$P_l = \text{REDUCE } (P_{l-1}).$$

Pyramids convert local image features into global features. The global information can be used to impose constraints on local operations. As a result hierarchical operations can be more efficient than operations performed at a single level of resolution.

2.3 The construction of a multiresolution image decomposition

A hierarchical decomposition of an image into a set of light and dark blobs can be obtained by successive application of size selective filters of various sizes. We noted before that size selective filters are equivalent to a difference of size limiting filters. As a result a pyramidal image decomposition can be computed as the difference between successive levels in a pyramidal image representation constructed with size limiting filters. Since these levels differ in sample density it is necessary to interpolate new values between the given samples in an image at a higher pyramid level before this image can be subtracted from the image residing at a lower pyramid level.

Interpolation can be achieved simply by the EXPAND operation:

$$P_{l,0} = P_l$$

and

$$P_{l,k} = \text{EXPAND}\ (P_{l,k-1})$$

where $P_{l,k}$ represents the image obtained by k successive applications of the EXPAND operation to P_l.

Let P be a pyramid constructed with size limiting filters. An error pyramid E is family of error images in which each image contains only details within a restricted range of sizes (i.e. a size selectively filtered set of images or a sieve). As noted before E can be computed as the difference between successive levels of P:

$$E(P)_l = P_l - \text{EXPAND}\ (P_{l+1}) \quad \text{for all } l \leq N$$

and

$$E(P)_N = P_N.$$

Because we are primarily interested in merging images for visual display we demand that visually important details of the component images must be preserved in the resulting composite image. It is a well known fact that the human visual system is sensitive to local luminance contrast. If an image fusion scheme is to preserve visually important details it must exploit this fact.

We now present an image decomposition scheme that is based on local luminance contrast (Toet, 1989a). This scheme computes the ratio of the low-pass images at successive levels of the Gaussian

pyramid. A multiresolution image decomposition based on local luminance contrast is obtained by computing a sequence of ratio images $R(P)_l$ defined by

$$R(P)_l = P_l / \text{EXPAND}\ (P_{l+1}) \quad 0 \le l \le N-1$$

and

$$R_N = P_N.$$

Thus every level R_l is a ratio of two successive levels in the size limiting pyramid P.

Luminance contrast C is defined as

$$C = \frac{L - L_b}{L_b}$$

$$= \frac{L}{L_b} - 1$$

where L denotes the luminance at a certain location in the image plane, L_b represents the luminance of the local background. Let $I_l(i,j) = 1$ for all i,j,l represent the *unit pyramid*. When C_l is defined as

$$C_l = P_l / \text{EXPAND}\ (P_{l+1}) - I_l$$

we have

$$R_l = C_l + I_l.$$

Therefore we will refer to the sequence R_l as the *contrast pyramid*.

The contrast pyramid is a complete representation of the original image. P_0 can be recovered exactly by reversing the steps used in the construction of the pyramid:

$$P_N = R_N$$

and

$$P_l = R_l \cdot \text{EXPAND}\ (P_{l+1}) \quad 0 \le l \le N-1.$$

This property of the contrast pyramid will be essential for the image merging scheme described below.

2.4 Linear filters

The application of linear filters in the construction of a size limited pyramid results in a sequence of images in which each is a low-pass filtered and subsampled copy of its predecessor.

A popular low-pass filter is the convolution with a Gaussian kernel. The application of this type of filter in the pyramid construction results in the well known Gaussian pyramid (Burt, 1984). In this case the filtering and sampling operations can be combined into a single operation by computing the Gaussian weighted average.

Let array P_0 contain the original image. This array represents the bottom or zero level of the pyramid structure. Each node of pyramid level l ($1 \leq l \leq N$ where N is the index of the top level of the pyramid) is obtained as a Gaussian weighted average of the nodes at level l-1 that are positioned within a 5 x 5 window centered on that node.

The REDUCE operation is defined by

$$P_l(i,j) = \sum_{m,n=-2}^{2} w(m,n)\, P_{l-1}(2i + m,\ 2j + n).$$

The weighting function $w(m,n)$ is separable:

$$w(m,n) = w'(m)w'(n),$$

with $w'(0) = a$, $w'(1) = w'(-1) = 0.5$, $w'(2) = w'(-2) = 0.5\ a$

A typical value of a is 0.4.

The corresponding EXPAND operation is given by

$$P_{l,k}(i,j) = 4 \cdot \sum_{m,n=-2}^{2} w(m,n)\, P_{l-1}\left(\frac{i+m}{2}, \frac{j+n}{2}\right),$$

where only integer coordinates contribute to the sum.

The error pyramid corresponding to the Gaussian pyramid is known in the literature as the DOG (Difference of Gaussians; e.g. Burton *et al.*, 1986), DOLP (Difference of Low Pass; e.g. Crowley and Parker, 1984) or Laplacian pyramid (e.g. Burt and Adelson, 1983).

2.5 Morphological filters

Multiresolution structural image representation and decomposition schemes typically apply linear (low- or band-pass) filters with progressively increasing spatial extent to generate a sequence of images with progressively decreasing resolution (e.g. Burt, 1983; Burt and Adelson, 1983; Burton *et al.*, 1986; Crowley and Parker, 1984). Linear filter techniques alter the object intensities and therefore the estimated location of their contours. As a result decomposition schemes based on these techniques are of limited applicability to tasks involving precise measurement of object size and shape.

Mathematical morphology examines the geometrical structure of an image by probing its microstructure with certain elementary forms (Serra, 1982). These so-called structuring elements are examined in the manner in which they fit into a set or the complement of a set. Thus the analysis is geometric in character. Moreover, it approaches image processing from the vantage point of human perception. The intent is to derive quantative measures of natural perceptual categories and thereby exploit whatever inherent congruences exist between image structure and ordinary human recognition.

Morphological filters remove image noise without adding a grayscale bias. They are therefore well suited for shape estimation.

We adopted alternating sequential morphological filters for the construction of the size limiting pyramid (Serra, 1988). Alternating sequential filters remove details of the image (fore- and background) which are small relative to the structuring element. Using classical terminology we can denote these alternating sequential filters as morphological low-pass filters. The filters are low-pass because it is high-frequency fluctuation between the set of pixel values and its complement which is attenuated in the output image. We used a square and flat (i.e. of uniform grayvalue or "brick-like") structuring element with an initial size of 5 x 5. Its size was doubled for each layer in the stack.

Morphological sampling does not allow a reconstruction whose positional accuracy is better than the radius of the circumscribing disk of the structuring element used in the reconstruction process, in contrast to the sampling reconstruction process in linear signal processing from which only frequencies below the Nyquist frequency can be reconstructed. In the discrete case this positional uncertainty may become fairly large (Toet, 1989b). To eliminate positional errors in the EXPAND process we dropped the sampling scheme altogether. This obviates the need for a reconstruction step.

In this case we end up with a stack of images with the same resolution but progressively diminishing structural content. The corresponding families of error and ratio images are easily obtained by respectively subtracting and dividing adjacent stack layers. A more extensive account of this method will be given in a future report (Toet, 1989c).

2.6 The image fusion scheme

The image merging scheme can be cast into a three step procedure. First a contrast pyramid is constructed for each of the source images. We assume that the different source images are in register and have the same dimensions. The latter restriction is not very serious as it can be shown that the method can also be applied to images with different definition regions as long as there is a common overlap. Second, a contrast pyramid is constructed for the composite image by selecting values from corresponding nodes in the component error pyramids. The actual selection rule will depend on the application and may be based on individual node values or on masks or confidence estimates. For example, in case of the fusion of two input images A and B into a single output image C, and maximum absolute local luminance contrast as a selection criterion we have (for all i,j,l):

$$R(C)_l(i,j) = R(A)_l(i,j) \quad \text{if } ||R(A)_l(i,j)-1|| > ||R(B)_l(i,j)-1||$$

and

$$R(B)_l(i,j) \quad \text{otherwise}$$

where R(A), R(B) represent the ratio or contrast pyramids for the two source images and R(C) represents the ratio pyramid for the fused output image. Finally, the composite image is recovered from its ratio pyramid representation through the EXPAND and multiply reconstruction procedure.

3 EXAMPLES

3.1 Simulation of binocular perception

This section illustrates the image fusion scheme by applying it on artificial input images with a low amount of detail. The results show a certain analogy of the pyramid image merging scheme and the human visual system's ability to accumulate information from two monocular images into a perceptual whole.

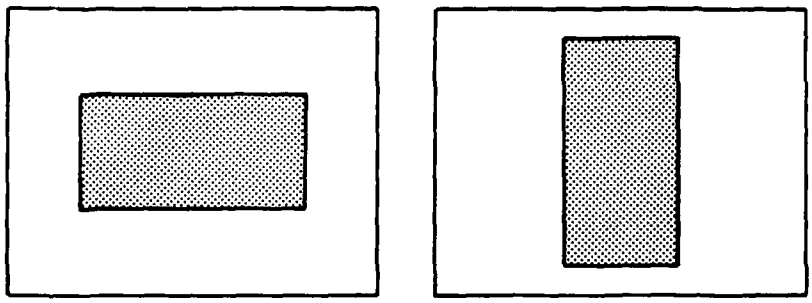

Fig. 2 Two test images.

Suppose a horizontal gray bar (Fig. 2a) is presented to the left eye of a human subject and a similar vertical bar (Fig. 2b) is simultaneously presented to the right eye of the subject. The resulting binocular percept is very similar to the result of the merging scheme shown in the lower right corner of Fig. 3. This result was obtained as follows. First the left and right bar images (denoted by L and R, respectively) are separately encoded as Gaussian ratio pyramids R(L) and R(R), respectively. A third so-called binocular pyramid R(B) is constructed by selecting the node with maximum absolute contrast value from the corresponding nodes in the left and right monocular pyramids. That is, for all i,j and l we have

$$R(B)_l(i,j) = R(L)_l(i,j) \quad \text{if } ||R(L)_l(i,j)-1|| > ||R(R)_l(i,j)-1||$$

$$= R(R)_l(i,j) \quad \text{otherwise.}$$

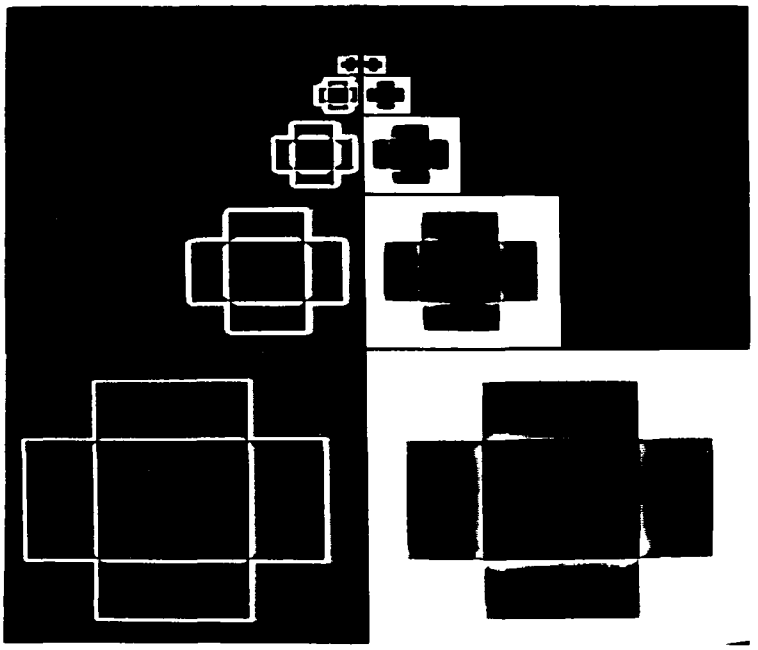

Fig. 3 The composite Gaussian ratio pyramid (left column) of Fig. 2 and the reconstructed composite Gaussian pyramid (right column). The lower level of the composite pyramid represents the final result of the image fusion scheme.

The binocular pyramid R(B) is shown in the left column of Fig. 3. The right column of this figure shows the reconstructed Gaussian pyramid for the binocular percept. The bottom layer of this pyramid represents the actual binocular combination. Notice that the edges of both images are preserved in the binocular image even though they are not specifically encoded. The shape and extent of the halo surrounding the central square represents the cumulative contribution of Laplacian filters of many spatial scales.

Fig. 4a and 4b show two gray bars containing smaller bars in their interior. The absolute luminance contrast between the small and large bars in Fig. 4 was chosen to be larger for the small bright bar in Fig. 4a than for the small dark bar in Fig. 4b. However, the absolute gray value difference between the small and large bars was smaller for the bright bar in Fig. 4a than for the dark bar in Fig. 4b. The human visual system is only sensitive to local luminance

contrasts. Thus, if Figs. 4a and 4b are presented simultaneously to respectively the left and right eye of a subject, the small bright bar of Fig. 4a will dominate the perception of the small dark bar of Fig. 4b in the resulting binocular percept. Fig. 5 and 6 show respectively the Gaussian or size limiting pyramid and the contrast or size selective pyramid of both images from Fig. 4. The composite contrast pyramid obtained by the maximum absolute contrast selection rule is shown on the left in Fig. 7. The reconstructed composite Gaussian pyramid is shown on the right in Fig. 7. Notice that the bright bar dominates the dark bar in this reconstruction in agreement with the representation by the human visual system.

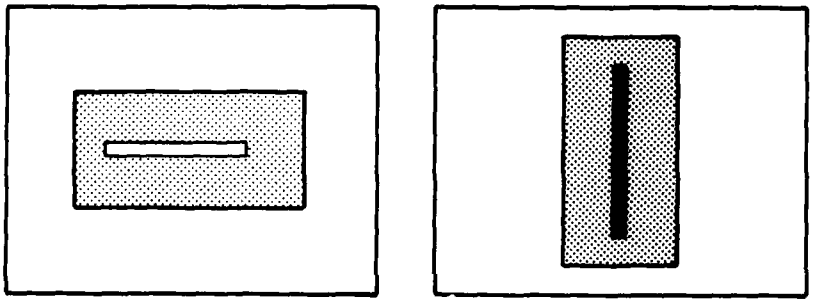

Fig. 4 Two test images.

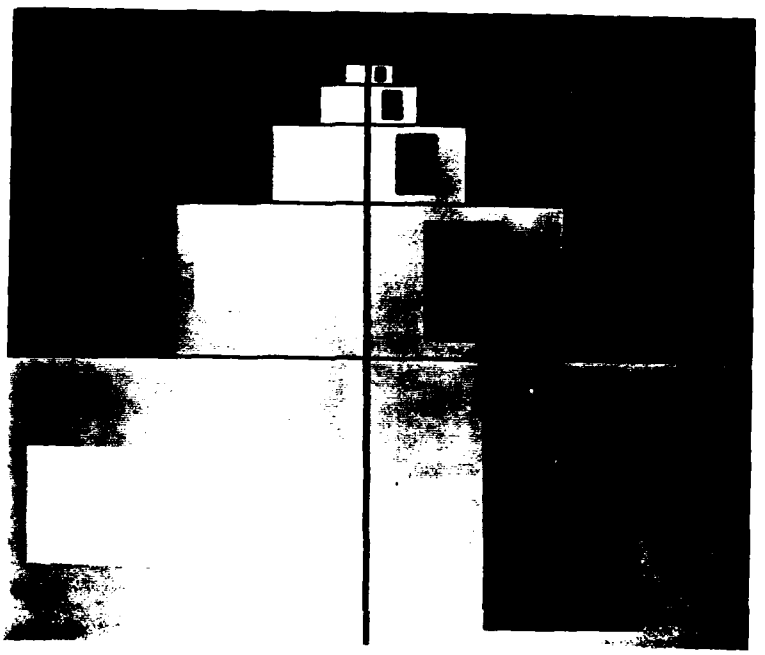

Fig. 5 The Gaussian pyramids for the images of Fig. 4.

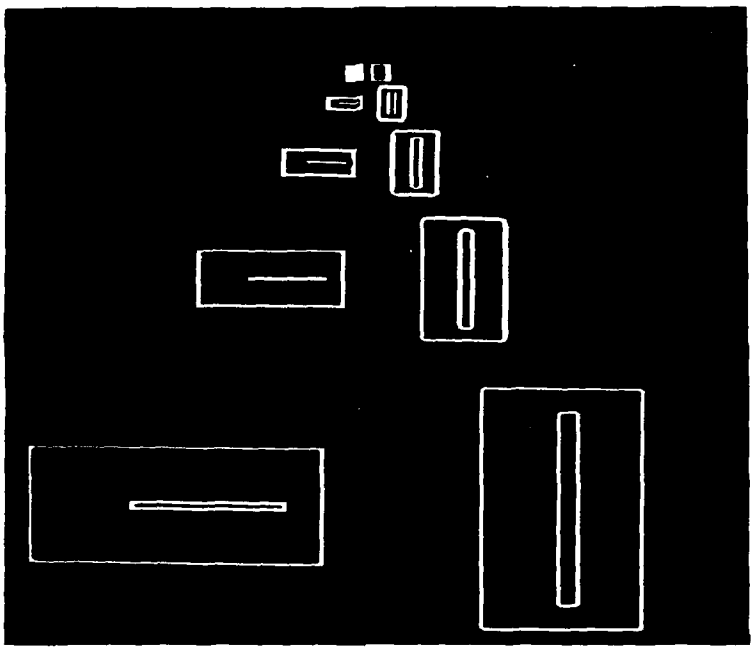

Fig. 6 The Gaussian ratio pyramids for the images of Fig. 4 constructed from Fig. 5.

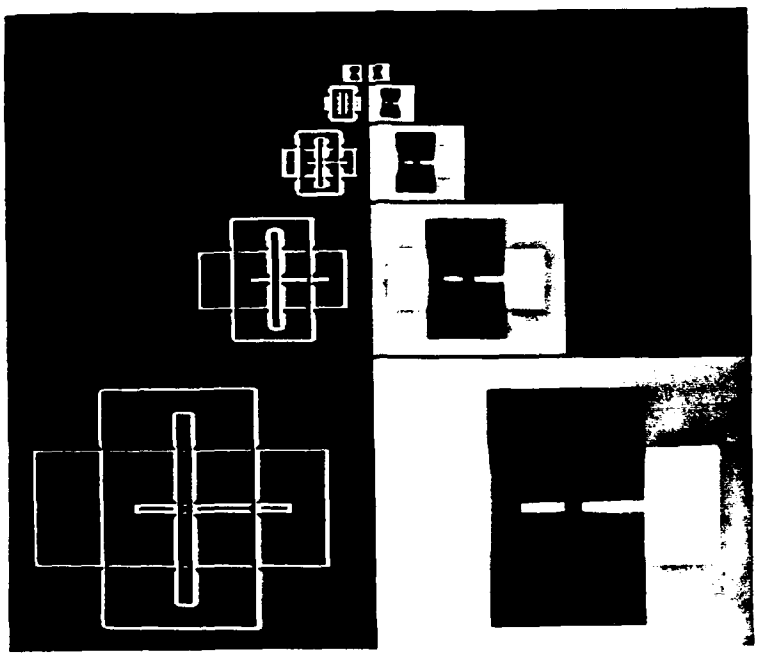

Fig. 7 The composite Gaussian ratio pyramid (left column) and the reconstructed composite Gaussian pyramid (right column) for the images of Fig. 4.

3.2 Fusion of thermal and visual images

In this section we present some results of the application of the image fusion scheme on real images. First we simultaneously recorded spatially registered CCD and FLIR images on video tape. The images were thereafter digitized and brought in register. Finally we digitally merged corresponding images using the hierarchical contrast merging scheme.

Aim of the experiment

Integration of visual (CCD) and thermal (FLIR) images can produce information that cannot be obtained by viewing the sensor outputs separately and consecutively. In defense applications for example, (details of) targets that are hard to detect in a visual image (that have low visual contrast) can sometimes easily be noticed in a thermal image. Incomplete representation of targets in thermal images

may result from large temperature gradients within these objects. Also, the exact location of targets in FLIR images may be hard to asses when the background has low thermal contrast. The increased information content of integrated FLIR and CCD images is expected to improve observer performance for a range of different tasks, e.g. the control of remotely piloted vehicles, driving in hostile environments and surveillance.

Image acquisition

Fig. 8 shows a schematic drawing of the experimental setup used to record the CCD and FLIR images (cf. Alferdinck, 1988). The CCD and FLIR cameras are directed along the same optical axis. This is done by using a slanting germanium mirror. Germanium transmits thermal radiation while reflecting visible light. Spatial image registration was obtained by creating a common Cartesian coordinate grid for both image modalities. This was done by placing 9 light bulbs in the scene. The bulbs were attached to 3 vertically erected equidistant poles. They were clearly visible in both image modalities and small enough to provide well-defined reference points. For spatial calibration of the recordings the CCD and FLIR images were displayed on the R and G channels of an RGB monitor. Spatial registration was obtained by *superimposing* the R and G grid points (i.e. the corresponding images of the light bulbs) through adjusting the magnification, tilt and direction of the CCD camera.

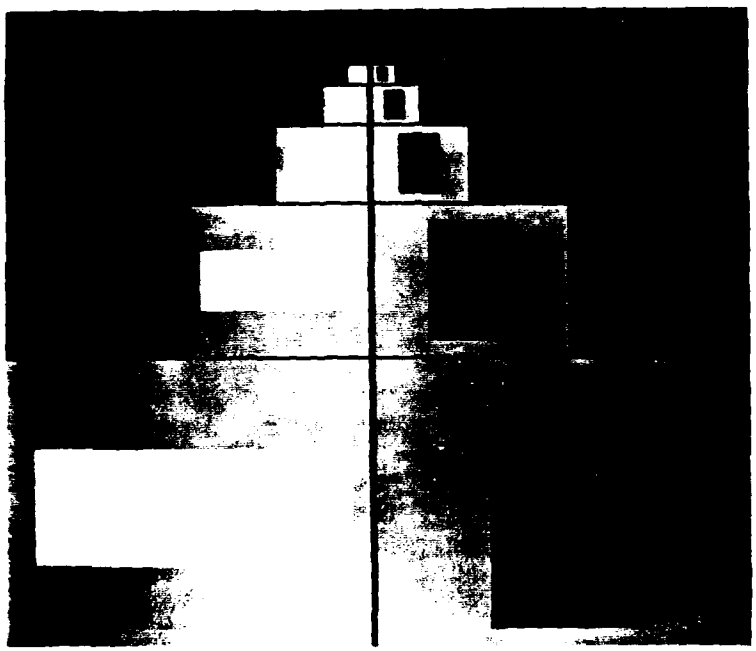

Fig. 8 Schematical representation of the experimental setup.

Prior to each recording session some shots of the reference coordinate grid were taken. After digitization of the recordings these shots were used to correct for small image distortions. This was done by affine warping transforms. The bulbs were removed prior to recording objects in the scene. The signals from both cameras were recorded on synchronized U-matic video taperecorders.

Image merging

Fig. 9 shows the CCD (Fig. 9a) and FLIR (Fig. 9b) images of a tank. In the CCD image it is hard to distinguish the front part of the gun barrel, the back part of the vehicle (containing the engine room) and the man hood located on top the tank. All these parts of the target have very low contrast in the visual image. However, in the thermal image these parts are clearly visible. The background, which is clearly visible in the CCD image, is nearly indistinguishable in the FLIR image.

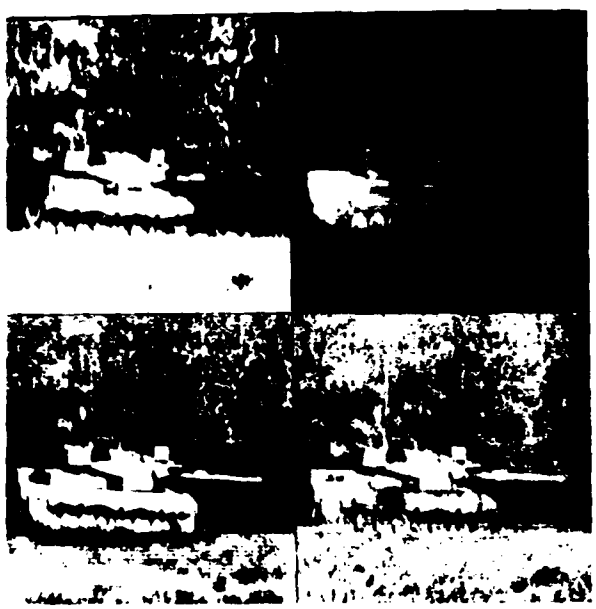

Fig. 9 Original CCD (a) and FLIR (b) images with the results of the linear (c) and morphological (d) merging schemes.

Fig. 10 shows the Gaussian ratio pyramids of the CCD (left column) and FLIR (right column) images from Fig. 9. Notice that the front part of the gun barrel, the back part of the vehicle and the man hood on top the tank are accentuated in the Gaussian ratio pyramid of the FLIR image and only very weakly represented in the Gaussian ratio pyramid of the CCD image. However, the background which is accentuated in the Gaussian ratio representation of the CCD image is very noisy in the FLIR image.

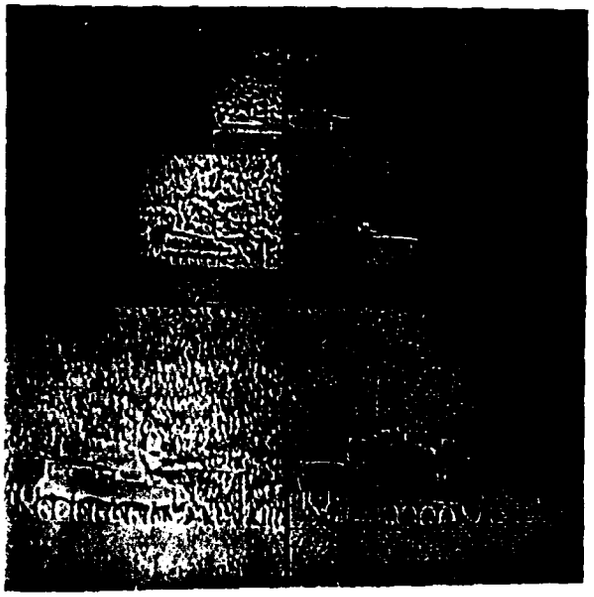

Fig. 10 The Gaussian ratio pyramids of the CCD (left column) and FLIR (right column) images from Fig. 9.

In this example we used the maximum absolute contrast node selection rule. Fig. 11 shows the composite Gaussian ratio pyramid (left column) and the reconstructed composite Gaussian ratio pyramid (right column). The bottom level of the composite Gaussian ratio pyramid represents the final result. This level is again shown in Fig. 9c, together with the original input images (Figs. 9a and b).

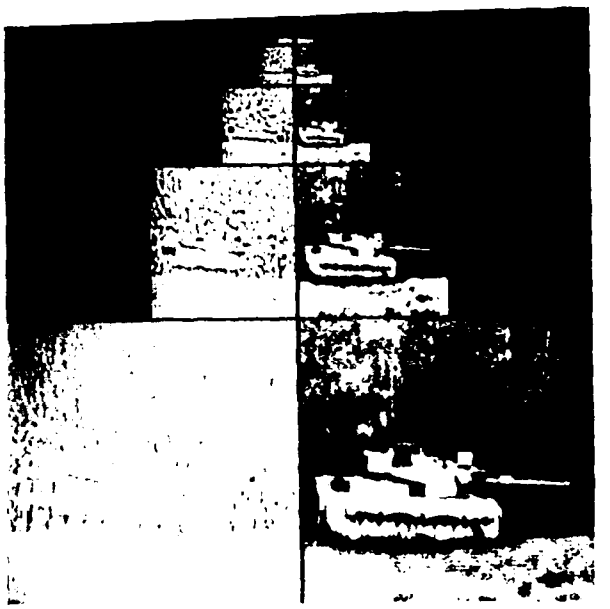

Fig. 11 The composite Gaussian ratio pyramid (left column) and the reconstructed composite Gaussian ratio pyramid (right column) of the CCD and FLIR images from Fig. 9a and b.

Fig. 12 and 13 show, respectively, the ratio stacks of the morphological filtered CCD and FLIR images from Fig. 9. Again we used the maximum absolute contrast node selection rule in the merging scheme.

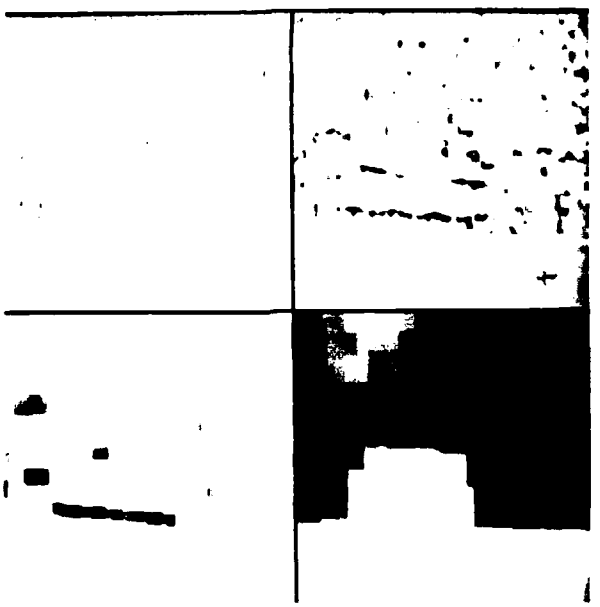

Fig. 12 The ratio stack of the morphological filtered CCD image from Fig. 9a.

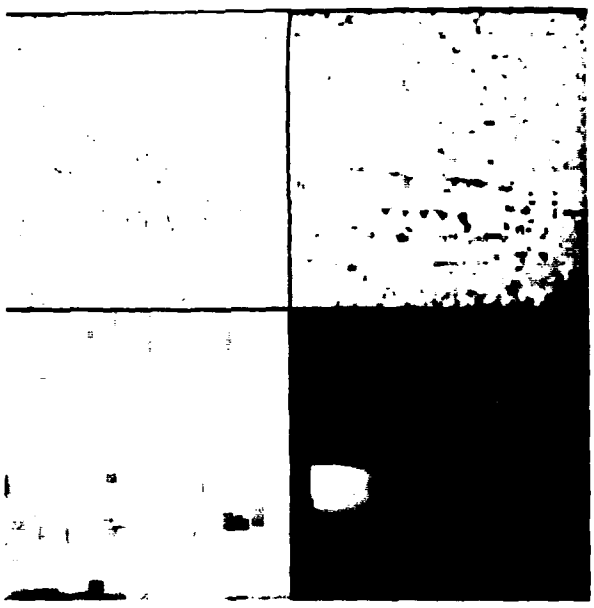

Fig. 13 The ratio stack of the morphological filtered FLIR image from Fig. 9b.

Fig. 14 shows the composite morphological ratio stack. The final result of the morphological merging process is shown in Fig. 9d, together with the original input images (Figs. 9a and b) and the result of the Gaussian decomposition scheme.

Fig. 14 The composite morphological ratio pyramid of the CCD and FLIR images from Fig. 9a and b.

Fig. 9 convincingly demonstrates that the fused images contain those details from both input images that have maximum local contrast. Notice that all (aforementioned) details that can only be obtained from a single image modality are clearly represented in the fused images.

Details in the composite image resulting from the morphological fusion scheme (Fig. 9d) are more pronounced (better articulated) than their counterparts in the result from the linear fusion scheme (Fig. 9c). This is a result of the fact that linear filters alter object intensities (blur image details) whereas morphological filters extract image details without adding a grayscale bias.

4 DISCUSSION AND CONCLUSIONS

In this report we presented an image fusion method intended for human observation. The method preserves details of high local luminance contrast.

First the input images are decomposed into sets of light and dark blobs on different levels of resolution. Thereafter the absolute contrast of blobs at corresponding locations and at corresponding levels of resolution are compared. The actual image fusion is done by selecting the blobs with maximum absolute luminance contrast. The fused image is reconstructed from the set of blobs or pattern primitives thus obtained. As a result perceptually important details (i.e. details with a relatively high local luminance contrast) of both images are preserved in the composite image.

In this report we also introduced a multiresolution image representation in which iterative morphological filters of many scales but identical shape serve as basis functions. A structural image decomposition obtained from this multiresolution representation differs from established techniques (like Gaussian blurring or Fourier decomposition) in that the primitives (i.e. image details) have a well defined location and size. Therefore, the resulting image description provides a useful basis for multiresolution shape analysis. Moreover, when vertical sided structuring elements are used in the filtering process the shape of object contours is well preserved across different levels of resolution.

The superiority of the morphological multiresolution image decomposition over a conventional linear multiresolution image decomposition is demonstrated by the results of the hierarchical image fusion scheme. The images produced by the morphological fusion scheme appear more crispy than the images produced by the corresponding linear fusion scheme. This is a result from the fact that morphological filters preserve local luminance contrast, whereas linear filters blur local luminance contrast.

The morphological image decomposition scheme is well suited for real-time (VLSI) implementation when binary set structuring elements are used. Because of the inherent congruences between the hierarchical morphological decomposition scheme and human visual perception the method appears well suited to eventual integration into artificial intelligent computer vision systems.

Image fusion based on luminance ratios only fails when the mean local background luminance of all of the input images is zero. A simple remedy is to add a constant background luminance to the composite image. This nonzero local background then serves as a pedestal for the contrast modulations.

The hierarchical image fusion scheme presented in this report is quite general and can be used to transfer any useful information from one image to the other. This can be done independently for each location in the scene and on every level of resolution. Thus, information present in the image obtained from sensor A can be used to filter information at corresponding locations in the image from sensor B. The choice of the filter operation will depend on the application (and can be anything from contrast enhancement to smoothing and thresholding). The method can be used to merge images from a variety of sensing modalities prior to display.

The fusion of data from different sensing modalities is complicated by the fact that the information may be represented in a different format and reference frame. Merging the information from two sensory systems is equivalent to mapping the topology of the sensory systems on a common topological space which is isomorphic to the environment. This requires an abstract algebraical representation of the sensor data which represents the topology of the object space. Such a representation can be obtained by computing a (Delaunay) triangulation of the data set. The data can then be combined by merging the different triangulations. Presently we are engaged in the development of this approach. This study is expected to result in a formalism for fusing data from totally different (i.e. of different dimensions and even non-imaging) sensing modalities.

REFERENCES

Alferdinck, J.W.A.M. and Valeton, J.M. (1988). Thermal and visual light images of military vehicles on "de Vlasakkers". Report IZF 1988-11, TNO Institute for Perception, Soesterberg, The Netherlands.

Burt, P.J. (1984). The pyramid as a structure for efficient computation. In: Multiresolution image processing and analysis, Rosenfeld, A. (ed.), Springer Verlag, Berlin, pp. 6-35.

Burt, P.J. and Adelson, E.H. (1983). The Laplacian pyramid as a compact image code, IEEE Tr. on Comm., vol. 31, pp. 532-540.

Burt, P.J. and Adelson, E.H. (1985). Merging images through pattern decomposition. In: Applications of digital image processing VIII, Proc. SPIE 575, pp. 173-181.

Burt, P.J., Hong, T.H. and Rosenfeld, A. (1981). Segmentation and estimation of image region properties through cooperative hierarchical computation, IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-11, pp. 802-809.

Burton, G.J., Haig N.D. and Moorhead, I.R. (1986). A self-similar stack model for human and machine vision. Biol. Cybern. 53, pp. 397-403.

Crowley, J.L. and Parker, A.C. (1984). A representation for shape based on peaks and ridges in the difference of low-pass transform. IEEE Trans. Patt. An. Mach. Int., vol. PAMI-6, pp. 156-170.

Eklundh, J.O., Lansner, A. and Wessblad, R. (1986). Classification of multispectral images using associative nets. In: Proc. 8th Int. Conf. on Pat. Rec., IEEE Comp. Soc. Press, Washington DC.

Koenderink, J.J. (1984). The structure of images. Biol. Cybern. 50, pp. 363-370.

Koenderink, J.J. and Doorn, A.J. van (1978). Visual detection of spatial contrast; influence of location in the visual field, target extent and illuminance level. Biol. Cybern. 30, 157-167.

O, Y.L. and Toet, A. (1989) Mathematical morphology in hierarchical image representation. In: Proc. NATO ASI The formation, handling and evaluation of medical images. Springer Verlag, New York. In press.

Pearson, J.C., Gelfand, J.J., Sullivan, W.E., Peterson, R.M. and Spence, C.D. (1988). Neural network approach to sensory fusion. In: Proc. SPIE vol. 931, Sensor Fusion, pp. 103-108.

Rosenfeld, A. (ed.) (1984). Multiresolution image processing and analysis. Springer Verlag, New York.

Serra, J. (1982). Image analysis and mathematical morphology. Academic Press, New York.

Serra, J. (1988). Alternating sequential filters. In: Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances, pp.203-216. Academic Press, New York.

Toet, A., Koenderink, J.J., Zuidema, P. and Graaf, C.N. de (1984). Image analysis - topological methods. In: Information processing in medical imaging, DeConinck, F. (Ed.), Martinus Nijhof, The Hague, pp. 306-342.

Toet, A. (1989a). Image fusion by a ratio of low-pass pyramid. Pattern Rec. Letters. In press.

Toet, A. (1989b). A morphological pyramidal image decomposition. Pattern Rec. Letters. In press.

Toet, A. (1989c). Morphological multiresolution image representations. Internal TNO report.

REPORT DOCUMENTATION PAGE

1. DEFENCE REPORT NUMBER (MOD-NL)	2. RECIPIENT'S ACCESSION NUMBER	3. PERFORMING ORGANIZATION REPORT NUMBER
TD 89-3363		IZF 1989-19
4. PROJECT/TASK/WORK UNIT NO.	5. CONTRACT NUMBER	6. REPORT DATE
731.2	A87/D/149	July 7, 1989
7. NUMBER OF PAGES	8. NUMBER OF REFERENCES	9. TYPE OF REPORT AND DATES COVERED
34	19	Final

10. TITLE AND SUBTITLE

Image fusion through multi-resolution contrast decomposition

11. AUTHOR(S)

A. Toet

12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TNO Institute for Perception
Kampweg 5
3769 DE SOESTERBERG

13. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

TNO Division of National Defence Research
Koningin Marialaan 21
2595 GA DEN HAAG

14. SUPPLEMENTARY NOTES

15. ABSTRACT (MAXIMUM 200 WORDS, 1044 BYTE)

Integration of images from different sensing modalities can produce information that cannot be obtained by viewing the sensor outputs separately and consecutively. This report introduces a hierarchical image merging scheme based on multi-resolution contrast decomposition. The composite images produced by this scheme preserve those details from the input images that are most relevant to visual perception. As an example the method is used to merge parallel registered thermal and visual images. The examples show that the fused images present a more detailed representation of the depicted scene. Detection, recognition and search tasks may therefore benefit from this new image representation.

16. DESCRIPTORS — IDENTIFIERS

Image processing
Pattern recognition

17a. SECURITY CLASSIFICATION (OF REPORT)	17b. SECURITY CLASSIFICATION (OF PAGE)	17c. SECURITY CLASSIFICATION (OF ABSTRACT)
-	-	-

18. DISTRIBUTION/AVAILABILITY STATEMENT

Unlimited availability

17d. SECURITY CLASSIFICATION (OF TITLES)
-

VERZENDLIJST

1. Hoofddirecteur van de Hoofdgroep Defensieonderzoek TNO

2. Directie Wetenschappelijk Onderzoek en Ontwikkeling Defensie

3. Hoofd Wetenschappelijk Onderzoek KL
   Plv. Hoofd Wetenschappelijk Onderzoek KL

4,5. Hoofd Wetenschappelijk Onderzoek KLu

6. Hoofd Wetenschappelijk Onderzoek KM
   Plv. Hoofd Wetenschappelijk Onderzoek KM

7. Wnd. Directeur Militair Geneeskundige Diensten
   Cdre vliegerarts H.H.M. van den Biggelaar

8. Inspecteur Geneeskundige Dienst KL
   Brig.Gen.-arts B.C. Mels

9. Inspecteur Geneeskundige Dienst KLu
   Cdre J.Th. Versteeg

10. Inspecteur Geneeskundige Dienst Zeemacht
    Cdr-arts A.J. Noordhoek

11. Directeur Wetenschappelijk Onderzoek en Ontwikkeling,
    t.a.v. A.M.B. Vorstenbosch

12. Directie Materieel KLu, Afd. Wetenschappelijke Ondersteuning,
    t.a.v. Vdg R. ter Heide

13. Ir. J. Rogge, KMA, Breda

14,15,16. Hoofd van het Wetensch. en Techn. Doc.- en Inform. Centrum
    voor de Krijgsmacht

LEDEN WAARNEMINGS CONTACT COMMISSIE

17. Maj.Ir. W.C.M. Bouwmans
18. LTZAR1 F.D.J.R. Feunekes
19. Dr. N. Guns
20. Drs. C.W. Lamberts
21. Ir. P.H. van Overbeek
22. Drs. W. Pelt
23. Maj. dierenarts H.W. Poen
24. Drs. F.H.J.I. Rameckers
25. Prof.Ir. C. van Schooneveld
26. LKol.Drs. H.W. de Swart
27. Kol. vliegerarts B. Voorsluijs

Extra exemplaren van dit rapport kunnen worden aangevraagd door tussenkomst van de HWOs of de DWOO.
